REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden to Washington Headquarters Service, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188) Washington, DC 20503.
PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.

1. REPORT DATE *(DD-MM-YYYY)* 31-05-2001	2. REPORT ~~DATE~~ TYPE Technical	3. DATES COVERED *(From - To)* May 2000-April 2001

4. TITLE AND SUBTITLE Algorithms for Networks and Link Structured Data	5a. CONTRACT NUMBER
	5b. GRANT NUMBER N00014-99-1-0463
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S) Kleinberg, Jon	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Cornell University 4130 Upson Hall Ithaca, NY 14853	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) Office of Sponsored Programs Attn: Susan Aylward 123 Day Hall Ithaca, NY14853	10. SPONSOR/MONITOR'S ACRONYM(S) OSP
	11. SPONSORING/MONITORING AGENCY REPORT NUMBER

12. DISTRIBUTION AVAILABILITY STATEMENT

Approved for Public Release

20010529 017

13. SUPPLEMENTARY NOTES

14. ABSTRACT

My research is concerned with algorithms that exploit the combinatorial structure of networks and information. Algorithmic perspectives are crucial both at the level of large-scale networks such as the Internet, as well as the virtual networked enviroments such as the Web that they support. In both contexts, it is important to design scalable techniques for managing the complexity of these enviroments, as well as robust models that can be used to guide the development of new algorithms.

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF: a. REPORT	b. ABSTRACT	c. THIS PAGE	17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
				5	Tammy Howe
					19b. TELEPONE NUMBER *(Include area code)* 607-255-0957

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI-Std Z39-18

ONR Young Investigator Annual Report

Title: Algorithms for Networks and Link-Structured Data

PI Name: Jon Kleinberg

Address: Dept. of Computer Science,
Cornell University
Ithaca, NY 14853

Phone number: 607-255-7316

Fax Number: 607-255-4428

Email address: kleinber@cs.cornell.edu

Award Number: N00014-99-1-0463

Web site: http://www.cs.cornell.edu/home/kleinber

Long term goals:

My research is concerned with algorithms that exploit the combinatorial structure of networks and information. Algorithmic perspectives are crucial both at the level of large-scale networks such as the Internet, as well as the virtual networked environments such as the Web that they support. In both contexts, it is important to design scalable techniques for managing the complexity of these environments, as well as robust models that can be used to guide the development of new algorithms.

Objectives:

The goal of this work is the design of methods for optimization and resource allocation in networks; for disseminating information in large decentralized networks; and for identifying clusters and important resources in a networked information environment. Underlying this is the goal of new models for understanding the dynamics by which information is propagated in networks, and the processes by which networks such as the Web evolve over time.

Approach:

As discussed above, the research involves both models and algorithms. At the level of models, the work has incorporated extensions to traditional random graph models, in which we study the effect of evolution over time. It also focuses on networks in which the components have very limited global knowledge of the overall topology and state. Moreover, these networks must operate continuously over time, reacting to events as they occur. Despite this lack of knowledge of global state and future events, the goal is to produce global behavior that is efficient and fault-tolerant.
The algorithmic techniques come from the areas of discrete optimization and randomized algorithms. Randomization is a powerful tool in decentralized network settings; for example, it forms the basis of robust "gossip" protocols that can be used to effectively spread information without global coordination. The research has augmented traditional notions from discrete optimization with algorithms guaranteeing solutions that are globally fair in multi-user environments. Optimization techniques are also used to formulate clustering and data mining problems from a combinatorial perspective.

Results:

A number of my recent results have been related to the complex structure and dynamic behavior of large networks. One basic property of networks such as the Web is their very small average node-to-node distances. This has been referred to as a "small-world" property after a series of famous experiments by the social psychologist Stanley Milgram in the 1960's. Milgram's experiments --- in which individuals tried to forward a letter to a designated "target," through people they knew on a first-name basis --- established not only that distances in the social network of personal acquaintances are very small, but that individuals operating with purely local information are actually able to find short paths. This is a fundamentally algorithmic insight -- a form of collective navigation without global knowledge or coordination. In recent work I provided a model for understanding this phenomenon, and efficient navigation algorithms within this model. The model is a generalization of one proposed by Watts and Strogatz, in which "long-range" links are added uniformly at random to an underlying D-dimensional lattice. Such networks are known to have small diameter. I demonstrate that if the long-range links are correlated in a precise way with the structure of the underlying lattice, then a natural decentralized algorithm can find short paths; but if the long-range links are not appropriately correlated, then local navigation is not possible. The result thus provides insight into the structural properties of a network that are needed to make this type of local routing possible.

With Alan Demers and my student David Kempe, I have found an interesting application of these idea to the design of gossip protocols in distributed networks. Gossip protocols are mechanisms by which nodes in such networks exchange information in discrete time steps with communication partners chosen according to some underlying deterministic or randomized algorithm. In many settings --- consider a network of sensors, or a cluster of distributed computing hosts --- new information is generated at individual nodes, and is most ''interesting'' to nodes that are nearby. Thus, we propose distance-based propagation bounds as a performance measure for gossip protocols: a node at distance d from the origin of a new piece of information should be able to learn about this information with a delay that grows slowly with d, and is independent of the size of the network. We show that if nodes communicate randomly with one another using a probability distribution that falls off polynomially in the distance separating them, then such rapid propagation is possible. We then show how such a propagation guarantee can be used to design a simple and robust protocol for resource location in such networks.

With my student Amit Kumar, I have continued to study resource allocation problems in multi-user environments. In many optimization problems, one seeks to allocate a limited set of resources to a set of individuals with demands. Thus, such allocations can naturally be viewed as vectors, with one coordinate representing each individual. Motivated by work in network routing and bandwidth assignment, we consider the problem of producing solutions that simultaneously approximate all feasible allocations in a coordinate-wise sense. This is a very strong type of ''global'' approximation guarantee, based on fairness; we explore its consequences in a range of discrete optimization problems, including facility location, scheduling, and bandwidth assignment in networks. A fundamental issue --- one not encountered in the traditional design of approximation algorithms --- is that good approximations in this global sense need not exist for every problem instance; there is no a priori reason why there should be an allocation that simultaneously approximates all others. As a result, the existential questions concerning such good allocations lead to a new perspective on a number of basic problems in resource allocation, and on the structure of their feasible solutions.

With Christos Papadimitriou and Prabhakar Raghavan, I have been studying an optimization-based framework for clustering and data mining in which one has access to data from a large population. We consider the extent to which designing effective algorithms can be balanced with concerns about the privacy of each individual's data. This involves algorithms for ''auditing'' of queries to determine whether they inadvertently reveal sensitive information; and -- for settings in which individuals are motivated to share private information -- the development of methods to determine the fair value of such information.

Impact/Applications:

The work on Web structure is closely tied to my continuing study of link-based Web search, which has led to the development of improved Web search techniques. My work on gossip protocols has proceeded in a collaborative interaction with the groups of Ken Birman and Robbert van Renesse at Cornell, who are designing scalable fault-tolerant distributed systems.
My work with Amit Kumar on resource allocation in networks has involved interactions with groups at Lucent Bell Labs, who are currently investigating the feasibility of a number of these ideas in practice.

Transitions:

Above, I discussed possible transitions of some of these ideas to groups involved with the design of fault-tolerant systems and communication networks. The algorithmic work on the small-world phenomenon has also led to follow-up studies by Huberman and his group at HP, concerned with searching in peer-to-peer networks. Finally, my algorithms for link-based Web search have been incorporated into several search projects and tools. This includes two of the largest scientific research digital libraries: the ResearchIndex Project at NEC, and the Cora Project at Whizbang! Labs.

Related Projects:

A number of related research projects have been identified above. In the area of Web search and structure analysis, work is being done by groups that include the Clever Project at IBM Almaden, Broder et al. at AltaVista, Henzinger et al. at Google, Raghavan et al. at Verity, and Lawrence et al. at NEC Research. Algorithms for network optimization and routing are being investigated by groups at including several at Lucent Bell Labs and AT&T Research; Plotkin et al. at Stanford; Awerbuch et al. at Johns Hopkins; Upfal et al. at Brown; and Karp, Papadimitrou, and Shenker at Berkeley and ACIRI.

Publications:

''Are randomly grown graphs really random?", submitted for publication, 2001. (with D. S. Callaway, J. E. Hopcroft, M. E. J. Newman, and S. H. Strogatz).

''Spatial Gossip and Resource Location Protocols," to appear at the 33rd ACM Symposium on Theory of Computing, 2001. (with D. Kempe, A. Demers).

''Provisioning a Virtual Private Networks: A Network Design Problem for Multicommodity Flow," to appear at the 33rd ACM Symposium on Theory of Computing, 2001. (with A. Gupta, A. Kumar, R. Rastogi, B. Yener).

''On the Value of Private Information," to appear at the 8th Conf. on Theoretical Aspects of Rationality and Knowledge, 2001. (with C. Papadimitriou, P. Raghavan).

''Adversarial queuing theory," Journal of the ACM 48(1) 13-38 (2001) (with A. Borodin, P. Raghavan, M. Sudan, D. Williamson)

''Universal-stability results and performance bounds for greedy contention-resolution protocols," Journal of the ACM 48(1) 39-69 (2001) (with D.M. Andrews, B. Awerbuch, A. Fernandez, F.T. Leighton, Z. Liu).

''Navigation in a Small World," Nature 406(2000), 845.

''Detecting a Network Failure," Proc. 41st IEEE Symposium on Foundations of Computer Science, 2000, 231-239.

''Fairness Measures for Resource Allocation," Proc. 41st IEEE Symposium on Foundations of Computer Science, 2000, 75-85. (with A. Kumar).

''Allocating Bandwidth for Bursty Connections," SIAM J. Computing 30(1) 191-217 (2000) (with Y. Rabani, E. Tardos).

''Node-Disjoint Paths on the Mesh, and a New Trade-Off in VLSI Layout," SIAM J. Computing 29(4) 1321-1333 (2000) (with A. Aggarwal, D. Williamson).

''Auditing Boolean Attributes." Proc. 19th ACM Symposium on Principles of Database Systems (2000), 86-91 (with C.H. Papadimitriou and P. Raghavan).

''The Small-World Phenomenon: An Algorithmic Perspective." Proc. 32nd ACM Symposium on Theory of Computing (2000), 163-170.

''Connectivity and Inference Problems for Temporal Networks." Proc. 32nd ACM Symposium on Theory of Computing (2000), 504-513 (with D. Kempe and A. Kumar).

''Query Strategies for Priced Information." Proc. 32nd ACM Symposium on Theory of Computing (2000), 582-591 (with M. Charikar, R. Fagin, V. Guruswami, P. Raghavan, and A. Sahai).

''Random Walks with 'Back Buttons'." Proc. 32nd ACM Symposium on Theory of Computing (2000), 484-493 (with R. Fagin, A. Karlin, P. Raghavan, S. Rajagopalan, R. Rubinfeld, M. Sudan, and A. Tomkins).